Experiments in Anomalous Change Detection
with the Viareggio 2013 Trial Dataset

James Theiler,[1] Michal Kucer,[1,2] and Amanda Ziemann[1]

[1]Intelligence and Space Research, Los Alamos National Laboratory, Los Alamos, NM 87545
[2]Center for Imaging Science, Rochester Institute of Technology, Rochester, NY 14623

ABSTRACT

The "Viareggio 2013 Trial" is a hyperspectral dataset obtained from multiple overflights of the Italian city of Viareggio. Careful management of panels and vehicles in the scene enabled the development of valuable ground truth information. One pair of overflights occurred at different times on the same day, and another pair took place over different days. These data were used to compare and evaluate a variety of automated approaches for discovering anomalous changes. Co-registration of the images is acknowledged to be imprecise, so part of the challenge is to identify anomalous changes in a way that is robust to this misregistration.

Keywords: anomalous change detection, hyperspectral imagery, remote sensing, local co-registration adjustment, elliptically-contoured distribution

1. INTRODUCTION

Most things are more than one thing.
— Maxwell Brian

Like many things, change detection is more than one thing. For change delineation, the aim is to identify with some precision the location and extent of a known change (e.g., after a flood, which streets are underwater?). For anomalous change detection, the change is presumed rare but the nature of that change is not specified. Change detection becomes more complicated with more than two images,[1–4] but even with only two images, there are multiple scenarios. One of the images might be considered a "reference" against which changes are sought in the other image. Or vice versa. Or, the problem may be more symmetrical: changes are sought without selecting either of the images as a reference. For two images X and Y, we can use a shorthand for these three cases: X→Y, Y→X, and X↔Y.

A variety of algorithms for anomalous change detection have been developed, including chronochrome (CC),[5] covariance equalization (CE),[6] multivariate alternation detection (MAD),[7] and hyperbolic anomalous change detection (HACD).[8] See Acito *et al.*[9] for a survey from a Gaussian point of view, and Ref. [10] for a much broader survey. One can extend these algorithms to elliptically-contoured (EC) distributions,[11] and more adaptive machine learning-based methods have also been developed.[12–14] A general machine learning framework was proposed in Ref. [15]; this framework distinguishes between *pervasive differences* and *anomalous changes*, and treats them as two classes in a binary classification problem. A resampling scheme can be used to provide an anomalous change pixel for every pervasive difference pixel, which enables matched-pair machine learning[16–18] approaches to be taken. In this machine learning framework, kernel-based methods[19–22] have often proved useful.

2. DATASET

The Viareggio 2013 Trial[23, 24] contains hyperspectral images form an airborne collect, taken in Viareggio, Italy, in 2013. Extensive ground truth is provided along with the imagery data, and the dataset has proved useful for investigating algorithms for target detection[25, 26] and for change detection.[27, 28] The change detection data consists of three hyperspectral images, which for simplicity of exposition we will call A, B, and C. (The longer names for these images are D1F12H1, D1F12H2, and D2F22H2, respectively.[24]) Two of the images (A and B) are taken on the same day an hour apart and a third image (C) is taken on the following day. The images are 450×375 pixels with 511 spectral channels in the visible and near infrared (400-1000 nm). Change maps
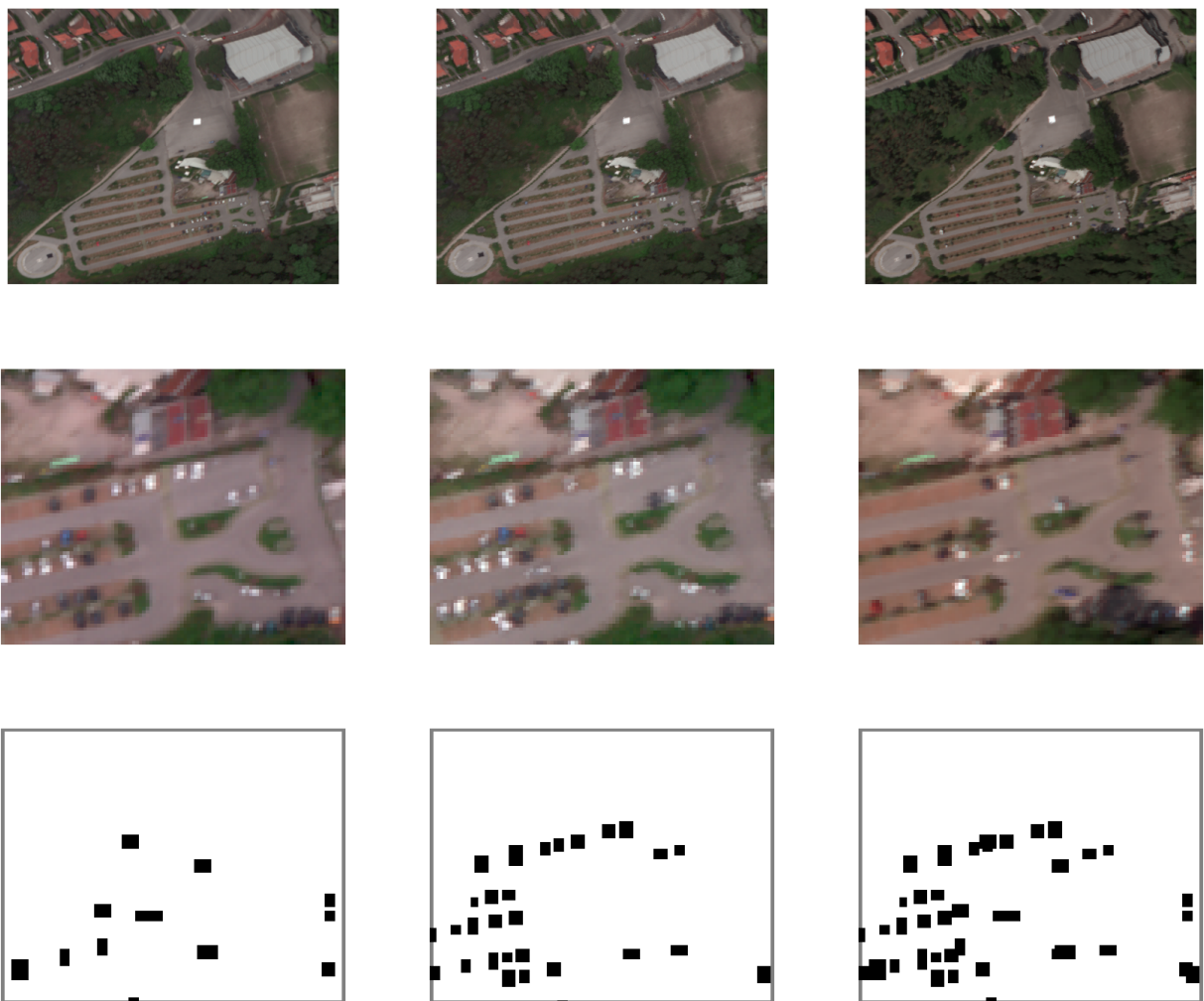
Figure 1. RGB images A, B, and C; obtained from channels 64, 32, and 22 of the spectrally binned imagery. Middle panels are close-ups of a particularly busy part of the parking lot, below and to the right of the center of the full image. Lower panels correspond to the inset-image change masks for A→C, C→A, and A↔C, respectively; each rectangle corresponds to a single object-level change.

are provided for the four asymmetric scenarios A→B, B→A, A→C, and C→A (where X→Y corresponds to "reference" image X and "test" image Y). Note that the A→B change map includes what is new in image B that is not present in A; the B→A map includes what was removed from B but was present in A. We will also consider the two symmetric change scenarios, A↔B and A↔C, with the change maps derived from an "or" operation applied to the two asymmetric change maps.

The Viareggio change masks are object based. Rather than define the specific pixels at which changes have occurred, each change is associated with a small rectangle inside of which a single change has occurred. The change can be anywhere from subpixel to a few pixels in extent, but is presumed to be contained entirely within the rectangle. If a change detection algorithm finds a change anywhere within the rectangle, it counts as a successful detection. If it finds multiple changed pixels within a single rectangle, it gets credit only for one detection. Detections observed outside the rectangles are false alarms, and these are counted on a pixel-wise basis.

3. EXPERIMENTS

For this study, we apply a variety of anomalous change detection algorithms to the Viareggio dataset. We measure performance in terms of false alarms and detections. The variation of detection rate (DR) versus false alarm rate (FAR), as a detection threshold is varied, leads to the widely used receiver-operator characteristic (ROC) curve. Since we are most concerned with the low false alarm rate regime of the ROC curve, we will focus on a single statistic: the false alarm rate at the threshold that provides a detection rate of one half (which we call FAR@DR=0.5).

We note that Wu *et al.*[28] recently performed a change detection study on the same imagery. That study considered both the A↔B and A↔C changes (though not the uni-directional changes: A→B, B→A, A→C, or C→A), and compared a fairly wide range of algorithms including RX, CC, HACD, subpixel-HACD,[29] and their own *slow feature analysis (SFA)*[30] – the latter two of which they found performed the best. Direct comparisons with our study, however, are hampered by some differences in how performance is measured. Although the authors report area under the ROC curve (AUC), which unfortunately puts most of the weight in the *high* false alarm rate regime of the curve, they also provide full ROC curves for some of their experiments, and based on these curves, FAR@DR=0.5 values appear to be 0.02 or higher. Also, it is not clear that their paper treats detections in an object-based way; if they require pixel-wise detections for the full area that defines where the anomalous changes are located, then they will report much lower performance than their algorithms actually achieved.

3.1 Comparison of pixel-wise algorithms

We begin with a comparison of three standard pixel-wise algorithms: straight anomaly detection on stacked pixels (RX), chronochrome (CC), and hyperbolic anomalous change detection (HACD). In later sections, we will extend these baseline implementations to include other improvements, and with those improvements in place, we can revisit this comparison.

With X and Y as the two images of interest, we have vector-valued pixels $\mathbf{x} \in \mathbb{R}^{d_x}$ and $\mathbf{y} \in \mathbb{R}^{d_y}$, with d_x and d_y the number of spectral channels in X and Y, respectively. Let $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ correspond to the mean values of $\mathbf{x}$ and $\mathbf{y}$, and let C_{xx} and C_{yy} be the covariance matrices associated with $\mathbf{x}$ and $\mathbf{y}$. That is: $C_{xx} = \left\langle (\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^\mathsf{T} \right\rangle$ and similarly for C_{yy}. Further, define a cross-covariance $C_{xy} = \left\langle (\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{x} - \boldsymbol{\mu}_x)^\mathsf{T} \right\rangle$. It is useful to define a "stacked" vector

$$\mathbf{z} = \left[\begin{array}{c} \mathbf{x} \\ \mathbf{y} \end{array} \right] \tag{1}$$

for which

$$\boldsymbol{\mu}_z = \left[\begin{array}{c} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{array} \right] \quad \text{and} \quad C_{zz} = \left[\begin{array}{cc} C_{xx} & C_{xy}^\mathsf{T} \\ C_{xy} & C_{yy} \end{array} \right]. \tag{2}$$

In terms of these expressions, we can define a family of quadratic covariance-based anomalous change detectors[8] that describe an "anomalousness of change" for a given pixel pair $(\mathbf{x}, \mathbf{y})$:

$$\mathcal{A}(\mathbf{x}, \mathbf{y}) = \xi_z - \beta_x \xi_x - \beta_y \xi_y \tag{3}$$

Algorithm	β_x	β_y	A↔B	A→B	B→A
RX	0	0	0.005301	0.005693	0.004870
CC	0	1	0.010089	0.015070	0.026578
CC	1	0	0.004568	0.002264	0.002085
HACD	1	1	0.001650	0.001860	0.003558

Algorithm	β_x	β_y	A↔C	A→C	C→A
RX	0	0	0.008265	0.003993	0.025474
CC	0	1	0.010038	0.021247	0.070802
CC	1	0	0.052787	0.001893	0.002458
HACD	1	1	0.004216	0.006904	0.003788

Table 1. Comparison of the baseline variants of standard pixel-wise change detection algorithms. The tabulated quantity is FAR@DR=0.5, that is: the false alarm rate at the threshold for which half of the targets are detected. Smaller values are better. While we show the baseline results here, a modified version of this table will be presented in Table 7 for more advanced implementations of these algorithms.

where

$$\xi_z = (\mathbf{z} - \boldsymbol{\mu}_z)^\mathsf{T} C_{zz}^{-1} (\mathbf{z} - \boldsymbol{\mu}_z), \tag{4}$$

$$\xi_x = (\mathbf{x} - \boldsymbol{\mu}_x)^\mathsf{T} C_{xx}^{-1} (\mathbf{x} - \boldsymbol{\mu}_x), \tag{5}$$

$$\xi_y = (\mathbf{y} - \boldsymbol{\mu}_y)^\mathsf{T} C_{yy}^{-1} (\mathbf{y} - \boldsymbol{\mu}_y). \tag{6}$$

Here, $\beta_x = \beta_y = 0$ corresponds to the ordinary RX-style anomalousness[31] of the stacked pixel $\mathbf{z}$. Choosing $\beta_x = \beta_y = 1$ leads to the hyperbolic anomalous change detector (HACD).[15] Both of these detectors are *symmetric* and make no distinction between X and Y in terms of which image is "reference" and which is changed with respect to that reference (in our shorthand notation, these algorithms are designed for the X↔Y scenario). The chronochrome (CC) detector[5] is an *asymmetric* detector, treating X as the "reference" image so that it is the X→Y scenario that is addressed, and here $\beta_x = 1$ and $\beta_y=0$. We can reverse this and take $\beta_x = 0$ and $\beta_y = 1$; this would correspond to the Y→X scenario.

Table 1 shows the results of these algorithms applied to the Viareggio dataset. For the symmetric changes (A↔B and A↔C), we find that the symmetric HACD algorithm achieves the fewest false alarms. For the directed (asymmetric) changes, the CC algorithm (in particular, the forward CC algorithm given by $\beta_x = 1$ and $\beta_y = 0$) achieves the lowest false alarm rate in three of the four cases, and is competitive with HACD in the fourth case (A→B). In what follows, unless otherwise noted, we will be using the HACD algorithm as the base pixel-wise anomalous change detector.

3.2 Local Co-Registration Adjustment (LCRA)

One of the most pernicious of the pervasive differences in multi-temporal imaging is misregistration. This occurs when correspondence between pixel location in an image and scene location on the ground is not consistent from image to image. Thus, the same pixel location can refer to different locations in the scene. It is clear that, for change detection purposes, it is desirable to co-register images as accurately as possible. But there will always be some residual misregistration error. The LCRA algorithm[32, 33] provides a way to improve change detection performance by making local adjustments. These local adjustments are made to minimize the measure of anomalous change at each pixel; although the adjustments may not be accurate in the direct sense of improving the actual co-registration, they can lead to more accurate change detection results. Using LCRA requires estimating the magnitude of residual misregistration error (RMRE),[34] though in this work, we consider a range of radii.

As defined in Sec. 3.1, consider the two images of interest X and Y with d_x and d_y spectral channels, respectively. At the image position indexed by (k, l), we write the vector-valued pixels as $\mathbf{x} = \mathrm{X}_{k,l} \in \mathbb{R}^{d_x}$ and $\mathbf{y} = \mathrm{Y}_{k,l} \in \mathbb{R}^{d_y}$. Our goal is to produce an "anomalousness" image A in which each scalar-valued pixel $A_{k,l}$

represents how anomalous the change is at the position (k,l). For pixel-wise ACD algorithms, we can write the anomalousness at (k,l) in terms of a function $\mathcal{A}(\mathbf{x},\mathbf{y})$ that depends only on the pixel values at (k,l):

$$A_{k,l} = \mathcal{A}(X_{k,l}, Y_{k,l}). \tag{7}$$

For changes X→Y, LCRA considers for each pixel in Y a window about the corresponding pixel in X, and chooses the pixel in this window that gives the lowest anomalousness.*

$$A_{k,l} = \min_{(m,n)\in W} \mathcal{A}(X_{k+m,l+n}, Y_{k,l}) \tag{8}$$

Here, $\mathcal{A}$ is the anomalousness function provided by the underlying pixel-wise ACD algorithm, and W is a set of integer pairs defining the optimization window. For changes X←Y, we run the window over the pixels in Y.

$$A_{k,l} = \min_{(m,n)\in W} \mathcal{A}(X_{k,l}, Y_{k+m,l+n}) \tag{9}$$

And for the symmetric case X↔Y, we have the SLCRA formula:

$$A_{k,l} = \max\left\{ \min_{(m,n)\in W} \mathcal{A}(X_{k,l}, Y_{k+m,l+n}), \min_{(m,n)\in W} \mathcal{A}(X_{k+m,l+m}, Y_{k,l}) \right\}. \tag{10}$$

The residual misregistration of the Viareggio images is evident from looking at the images themselves, and as Fig. 2 and Table 2 both show, the performance of anomalous change detection is substantially improved by implementing LCRA. For both same-day changes A↔B and next-day changes A↔C, we see improvements out to $r=5$ pixels, which provides a measure of the effective misregistration of the image pairs.

As Table 2 shows, for the symmetric change problems A↔B and A↔C, the symmetric SLCRA in Eq. (10) is generally better than the uni-directional LCRA algorithms in Eq. (8) and Eq. (9). For the directional changes, on the other hand, Table 3 shows that LCRA outperforms SLCRA.

We remark that Wu *et al.*[28] also considered LCRA in their Viareggio change detection study, but they only considered a radius of 1, and found that LCRA sometimes helped and sometimes did not. Given that we found improvements out to much larger radii, it seems that they may be missing the full utility of LCRA for this dataset.

Given the large radii at which LCRA minimized false alarm rates, a case can be made for more aggressive co-registration (e.g., as proposed by Zelinski *et al.*[35]) *before* deploying anomalous change detection. As long as there is some residual misregistration, then LCRA can still be helpful, but in general we expect better performance (and more efficient computation) when that residual misregistration is small.

3.2.1 Comparison of square and circular windows

The window used in LCRA need not be square, and since one does not expect misregistration to be aligned with pixel axes, there is an intuitive preference for circular-shaped windows. In Table 4, we compare the performance of LCRA using circular and square windows, and although we did not observe substantial differences between them in terms of performance at optimal window size, we do note that circular windows often achieve better performance at small radii, which is at least preferable from a computational standpoint.

3.3 Non-Maximal Suppression (NMS)

For non-maximal suppression, the anomalousness values are modified based on the local neighborhood (defined in terms of a window W, which for our experiments was a 5×5 square) of anomalousness values. Each pixel value is compared to the maximum pixel value in its neighborhood. If it is equal to that maximum, then it is retained; if it is smaller than that value, then it is suppressed:

$$A^*_{k,l} = \begin{cases} A_{k,l} & \text{if } A_{k,l} = \max_{(m,n)\in W} A_{k+m,l+n}, \\ A_{\min} & \text{otherwise} \end{cases} \tag{11}$$

---

*By the way, Eq. (2) and Fig. 3 in Ref. [32] have this backwards.

A↔B

window radius	SLCRA	SLCRA+NMS	LCRA	LCRA+NMS	rev-LCRA	rev-LCRA+NMS
0	0.001632	0.000524	0.001632	0.000524	0.001632	0.000524
1	0.000518	0.000208	0.001435	0.000578	0.003311	0.000977
2	0.000208	0.000107	0.002263	0.000929	0.002496	0.000995
3	0.000071	0.000030	0.000226	0.000149	0.002359	0.001662
4	0.000048	0.000006	0.000083	0.000036	0.002621	0.001644
5	0.000095	0.000012	0.000089	0.000006	0.003359	0.002347
6	0.000095	0.000012	0.000256	0.000006	0.005086	0.004515
7	0.000095	0.000012	0.000423	0.001418	0.008308	0.004217
8	0.000089	0.000012	0.000709	0.002835	0.008844	0.004401

A↔C

window radius	SLCRA	SLCRA+NMS	LCRA	LCRA+NMS	rev-LCRA	rev-LCRA+NMS
0	0.004031	0.000617	0.004031	0.000617	0.004031	0.000617
1	0.003623	0.000569	0.006887	0.001282	0.003743	0.000641
2	0.001940	0.000365	0.011631	0.002815	0.002605	0.000515
3	0.000826	0.000228	0.013086	0.005785	0.001683	0.000425
4	0.000593	0.000150	0.023741	0.013493	0.001210	0.000252
5	0.000479	0.000084	0.020303	0.017063	0.000982	0.000228
6	0.000491	0.000096	0.023303	0.025130	0.000886	0.000192
7	0.000437	0.000084	0.025142	0.025028	0.000934	0.000180
8	0.000443	0.000138	0.025405	0.024274	0.001030	0.000180

Table 2. False alarm rate at DR=0.5 for various LCRA window sizes, without and with non-maximal suppression (suppression diameter is 5 pixels). Here we exclusively analyze HACD, after using CCA to reduce the dimension to 20. The SLCRA data is plotted in Fig. 2
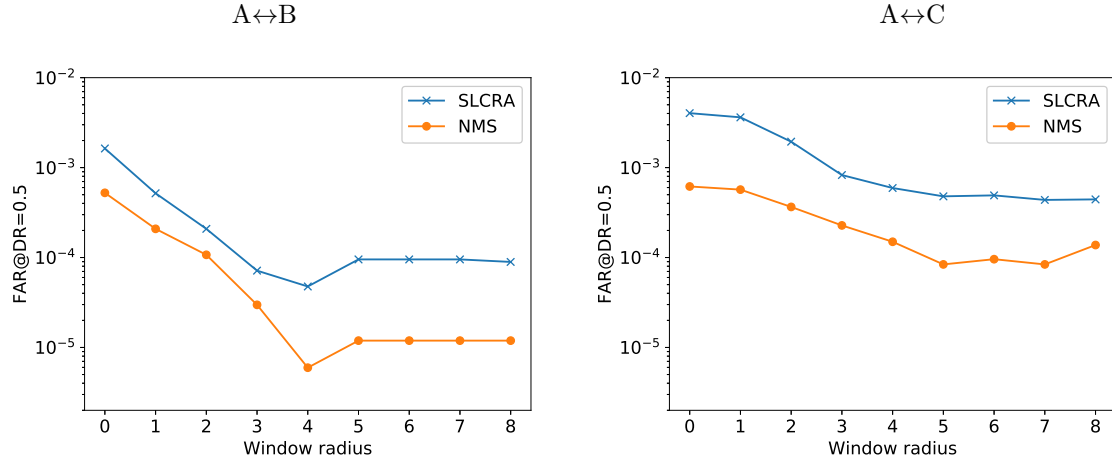

Figure 2. False alarm rate at DR=0.5 for various LCRA window sizes, without and with non-maximal suppression (suppression diameter is 5 pixels). Here we exclusively analyze HACD, after using CCA to reduce the dimension to 20. This same data is shown in Table 2.

A→B

radius	LCRA	LCRA+NMS	SLCRA	SLCRA+NMS
0	0.001842	0.000570	0.001842	0.000570
1	0.000398	0.000190	0.000725	0.000256
2	0.000101	0.000048	0.000475	0.000160
3	0.000018	0.000018	0.000368	0.000083
4	0.000012	0.000006	0.000428	0.000071
5	0.000006	0.000000	0.000458	0.000065
6	0.000006	0.000000	0.000452	0.000065
7	0.000006	0.000000	0.000464	0.000065
8	0.000012	0.000000	0.000517	0.000071

A→C

radius	LCRA	LCRA+NMS	SLCRA	SLCRA+NMS
0	0.006922	0.001012	0.006922	0.001012
1	0.003672	0.000673	0.006225	0.001024
2	0.001291	0.000357	0.003523	0.000655
3	0.000405	0.000107	0.002196	0.000470
4	0.000274	0.000060	0.001899	0.000357
5	0.000196	0.000036	0.001756	0.000321
6	0.000214	0.000036	0.001714	0.000315
7	0.000190	0.000036	0.001625	0.000298
8	0.000220	0.000048	0.001714	0.000423

B→A

radius	LCRA	LCRA+NMS	SLCRA	SLCRA+NMS
0	0.003356	0.000909	0.003356	0.000909
1	0.000689	0.000279	0.001461	0.000493
2	0.000119	0.000065	0.000499	0.000172
3	0.000036	0.000006	0.000303	0.000077
4	0.000036	0.000006	0.000285	0.000071
5	0.000089	0.000012	0.000517	0.000083
6	0.000089	0.000012	0.000511	0.000083
7	0.000089	0.000012	0.000505	0.000071
8	0.000089	0.000012	0.000487	0.000083

C→A

radius	LCRA	LCRA+NMS	SLCRA	SLCRA+NMS
0	0.003806	0.000591	0.003806	0.000591
1	0.001855	0.000364	0.003162	0.000543
2	0.000895	0.000215	0.001641	0.000346
3	0.000328	0.000101	0.000776	0.000215
4	0.000292	0.000054	0.000662	0.000161
5	0.000268	0.000036	0.000543	0.000137
6	0.000310	0.000054	0.000519	0.000149
7	0.000262	0.000048	0.000680	0.000167
8	0.000233	0.000054	0.000662	0.000167

Table 3. False alarm rate at DR=0.5 for various LCRA window sizes, without and with non-maximal suppression (suppression diameter is 5 pixels). Here we use HACD, after using CCA to reduce the dimension to 20. Note that for these asymmetric change problems, LCRA outperforms SLCRA.

A↔B: Circular window

radius of window	area of window	SLCRA	SLCRA+NMS
0	1	0.001632	0.000524
1	5	0.000518	0.000208
2	13	0.000208	0.000107
3	29	0.000071	0.000030
4	49	0.000048	0.000006
5	81	0.000095	0.000012
6	113	0.000095	0.000012
7	149	0.000095	0.000012
8	197	0.000089	0.000012

A↔B: Square window

radius of window	area of window	SLCRA	SLCRA+NMS
0	1	0.001632	0.000524
1	9	0.000423	0.000179
2	25	0.000101	0.000042
3	49	0.000071	0.000018
4	81	0.000101	0.000018
5	121	0.000095	0.000012
6	169	0.000089	0.000012
7	225	0.000089	0.000012
8	289	0.000113	0.000018

A↔C: Circular window

radius of window	area of window	SLCRA	SLCRA+NMS
0	1	0.004031	0.000617
1	5	0.003623	0.000569
2	13	0.001940	0.000365
3	29	0.000826	0.000228
4	49	0.000593	0.000150
5	81	0.000479	0.000084
6	113	0.000491	0.000096
7	149	0.000437	0.000084
8	197	0.000443	0.000138

A↔C: Square window

radius of window	area of window	SLCRA	SLCRA+NMS
0	1	0.004031	0.000617
1	9	0.002767	0.000497
2	25	0.001186	0.000293
3	49	0.000605	0.000150
4	81	0.000497	0.000102
5	121	0.000461	0.000090
6	169	0.000371	0.000096
7	225	0.000419	0.000138
8	289	0.000437	0.000132

Table 4. Comparison of circular to square windows (see Fig. 3) for SLCRA and SLCRA+NMS applied to the symmetrical change detection scenarios.
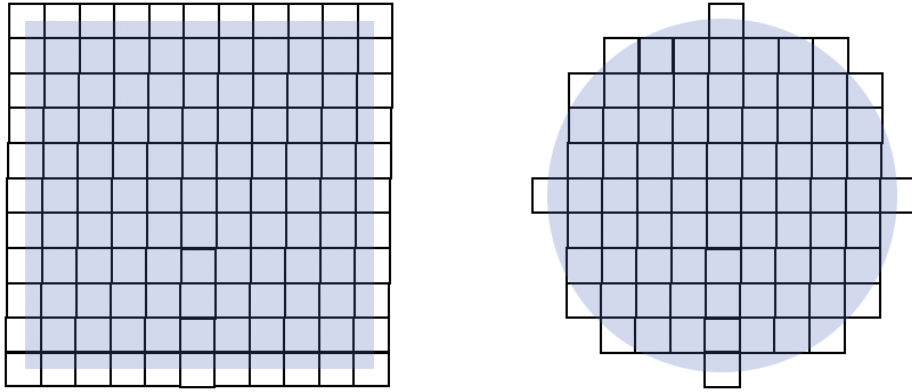

Figure 3. For a window of radius 5, a square window has an area of 121 pixels, while a circular window has an area of 81 pixels. (See Table 4 for a comparison of SLCRA performance with circular and square windows.)

where $A_{\min}$ is the minimum of all the anomaly values over the whole image. Since false alarms tend to occur in clumps, this approach suppresses all but the most anomalous pixel in the clump. Of course, if there were two actual changes in the same 5x5 window, one of them would be suppressed, and non-maximal suppression would be detrimental to performance. The choice of window size is thus driven by how far apart the actual changes are presumed to be.

We observe that NMS reduces false alarm rates often by a factor of over five. This is dramatic, at least on paper. One could quibble over whether it would actually help human analysts ultimately identify the interesting changes, however, because the analyst will also be inclined to discount clumps of false alarms, and will not necessarily investigate each one independently. But even in that case, if one is providing the analyst with, say, a "top ten" list of potentially interesting anomalous changes, NMS can be used to avoid putting anomalies from the same clump into that list.

3.4 Canonical Correlation Analysis (CCA)

CCA is a dimension reduction scheme that is similar in flavor to Principal Components Analysis (PCA), but where PCA identifies linear combinations of components that maximize the variance of a single vector-valued random variable, CCA instead finds linear combinations that maximize the correlation between pairs of vector-valued random variables.

We can illustrate the concept by considering just the first component in a CCA analysis. Choose unit vectors $\mathbf{a}$ and $\mathbf{b}$ so that corresponding pixels $\mathbf{x}$ and $\mathbf{y}$ from the two images, are transformed to scalar values $\mathbf{a}^\mathsf{T}\mathbf{x}$ and $\mathbf{b}^\mathsf{T}\mathbf{y}$. The idea is to choose these transformations so as to maximize the correlation of the scalars: $\langle [\mathbf{a}^\mathsf{T}(\mathbf{x}-\boldsymbol{\mu}_x)][\mathbf{b}^\mathsf{T}(\mathbf{y}-\boldsymbol{\mu}_y)] \rangle = \mathbf{a}^\mathsf{T} \langle (\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{y}-\boldsymbol{\mu}_y)^\mathsf{T} \rangle \mathbf{b}$, the solution of which can be expressed in terms of the left and right eigenvectors of the cross-covariance $\langle (\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{y}-\boldsymbol{\mu}_y)^\mathsf{T} \rangle$. The relationship of CCA to CE[6] and MAD[7] is described in Ref. [8]. (Note that the slow features in Slow Feature Analysis, as applied to pairs of images,[30] are essentially the canonical components of CCA.)

By finding linear combinations of spectral channels that maximize the correlation of two multispectral or hyperspectral images, and in particular that maximize this correlation *on average over the whole image*, the effect is to suppress pervasive differences between the images, and thereby to make the (rare) anomalous changes stand out in greater relief.

As Fig. 4 shows, CCA can be used to reduce the dimension considerably, and at virtually no cost in terms of change detection performance; in some cases, it can improve performance. As a rule, we recommend CCA as a preprocessing step in change detection, because even when it doesn't improve performance, it reduces computational cost. For most of the analysis in this manuscript, we used CCA to reduce the dimension to 20 (empirically chosen).

3.5 Preprocessing: de-striping, noise whitening, and spectral binning

As outlined in Fig. 7 of Ref. [24], various levels of preprocessing were applied to the original hyperspectral data in the Viareggio 2013 Trial dataset, leading to four variants of the image data being provided to users. The first stage is a de-striping (*ds*), and it is applied for all of the datasets. A second stage is a noise whitening (*nw*) step. Finally, both the *ds* and *nw* variants of the data are provided both at full (*ful*) spectral resolution (511 channels), and with the spectral data binned (*bin*) into 127 channels. Thus the four variants are: *ds-ful*, *nw-ful*, *ds-bin*, and *nw-bin*. Although these may result in substantial differences for direct target detection, we observe in Table 5 that they have relatively little effect on anomalous change detection performance. For the day-apart A↔C changes, the noise whitening does seems to help somewhat, and an improvement is also observed in the hour-apart A↔B changes, but it is very small. Since we find that using CCA to reduce dimension (even as low as 20 channels) is often helpful (and certainly speeds up the computation), it is not surprising that the spectrally binned data, which reduces the 511 channels to 127 channels, does not suffer observable information loss in the context of anomalous change detection.
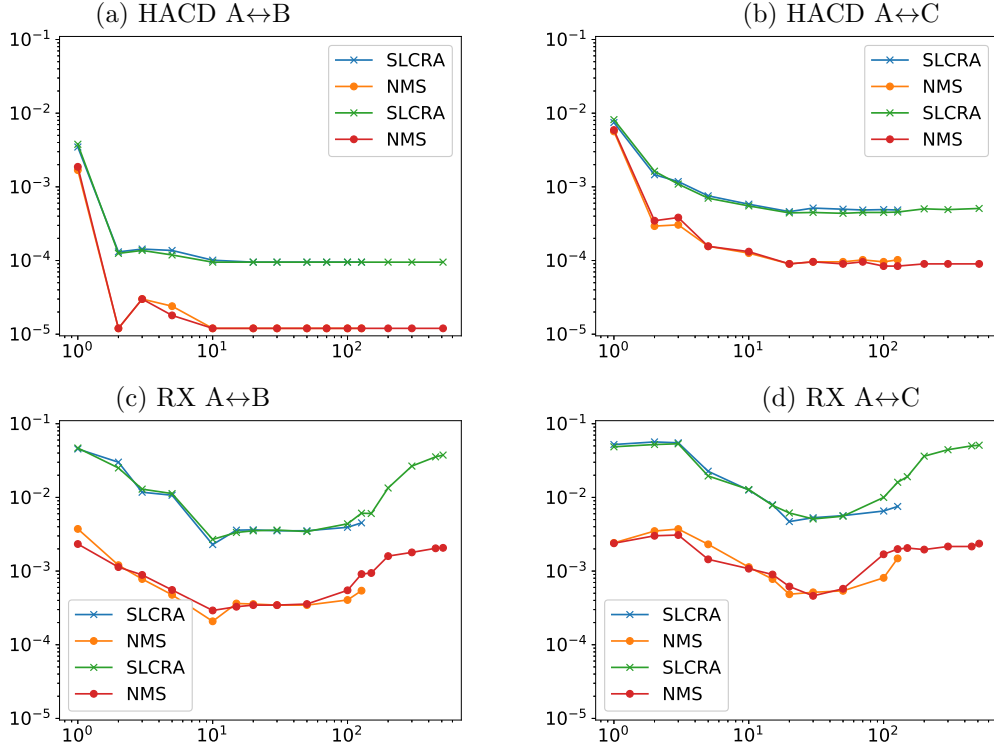
Figure 4. Canonical Correlation Analysis (CCA) is used to reduce the dimension of the hyperspectral data before change detection is applied. The false alarm rate (vertical axis) is plotted against the dimension (horizontal axis) for both the binned (127 channels) and the full (511 channels) hyperspectral images. It is evident in (a,b) that reducing the dimension to 20 leaves the HACD performance intact, and in (c,d) we see that dimension reduction substantially improves the performance, reducing the false alarm rate by roughly a factor of ten.

CCA=20		A↔B		A↔C	
spectral channels	noise whitening	SLCRA	SLCRA+NMS	SLCRA	SLCRA+NMS
127	-	0.000048	0.000006	0.000593	0.000150
127	yes	0.000071	0.000012	0.000329	0.000102
511	-	0.000042	0.000006	0.000479	0.000120
511	yes	0.000066	0.000012	0.000347	0.000090

w/o CCA		A↔B		A↔C	
spectral channels	noise whitening	SLCRA	SLCRA+NMS	SLCRA	SLCRA+NMS
127	-	0.000048	0.000006	0.000611	0.000138
127	yes	0.000071	0.000012	0.000365	0.000102
511	-	0.000042	0.000006	0.000485	0.000114
511	yes	0.000048	0.000012	0.000359	0.000072

Table 5. Spectral binning and noise whitening, with and without canonical correlation analysis (CCA) dimension reduction; here the window radius is 4, and the top line corresponds to the $w = 4$ line in Table 2.

	A↔B			A↔C	
ν	SLCRA	SLCRA+NMS	ν	SLCRA	SLCRA+NMS
∞	0.000095	0.000012	∞	0.000479	0.000084
50	0.000131	0.000024	50	0.000862	0.000186
20	0.000131	0.000024	20	0.001090	0.000216
10	0.000149	0.000024	10	0.000982	0.000198
5	0.000155	0.000024	5	0.001108	0.000228
5.59*	0.000155	0.000024	4.91*	0.001120	0.000228

Table 6. For elliptically-contoured HACD, the parameter ν characterizes the multivariate-t distribution. The asterisked values correspond to the value of ν estimated from the data, using the moment estimation formula.[11] Although improved performance has previously been observed using EC-HACD, and although theory supports using a distribution (multivariate-t) that is more adapted to the data than a Gaussian, we observe here that our best performance is obtained using the Gaussian-based HACD algorithm (*i.e.*, effectively using $\nu = \infty$).

Algorithm	β_x	β_y	A↔B		A→B		B→A	
			SLCRA	SLCRA+NMS	LCRA	LCRA+NMS	LCRA	LCRA+NMS
RX	0	0	0.003627	0.000357	0.002585	0.000327	0.002340	0.000285
CC	0	1	0.002531	0.000393	0.001129	0.000196	0.001176	0.000220
CC	1	0	0.001977	0.000339	0.000737	0.000172	0.001520	0.000285
HACD	1	1	0.000095	0.000012	0.000006	0.000000	0.000089	0.000012

Algorithm	β_x	β_y	A↔C		A→C		C→A	
			SLCRA	SLCRA+NMS	LCRA	LCRA+NMS	LCRA	LCRA+NMS
RX	0	0	0.004695	0.000485	0.002410	0.000250	0.004134	0.000495
CC	0	1	0.002318	0.000347	0.002369	0.000298	0.002726	0.000304
CC	1	0	0.006887	0.000964	0.001428	0.000214	0.001503	0.000215
HACD	1	1	0.000479	0.000084	0.000196	0.000036	0.000268	0.000036

Table 7. Comparison of algorithms: straight anomaly detection on stacked pixels (RX), chronochrome (CC), and hyperbolic anomalous change detection (HACD). In contrast to Table 1, we used LCRA, NMS, and CCA as part of the change detection algorithms. Here, SLCRA is used for the symmetrical change scenarios (↔), and LCRA for the asymmetric change scenarios (→). Circular windows of radius 5 were used for all the LCRA runs, and a square 5×5 window was used for the NMS. CCA was used to reduce the dimension to 20 before doing the change detection. The ds-bin (destriped, spectrally binned) variant of the imagery was used. We used the Gaussian variants of RX, CC, and HACD. With these various improvements, the false alarm rates reported here are much smaller than the corresponding values for the no-frills algorithms in Table 1. A qualitative difference we observe is that for these improved change detections, HACD outperforms CC and RX not only on the symmetric but on the asymmetric changes as well.

3.6 Elliptically-contoured change detection

If instead of assuming the background is Gaussian, one assumes that it is multivariate t-distributed with parameter ν, then the expression for anomalousness of change replaces Eq. (3) with

$$\mathcal{A}(\mathbf{x}, \mathbf{y}) = F_\nu(\xi_z) - \beta_x F_\nu(\xi_x) - \beta_y F_\nu(\xi_y) \tag{12}$$

where

$$F_\nu(\xi) = (d + \nu) \log\left(1 + \frac{\xi}{\nu - 2}\right), \tag{13}$$

and d is the number of spectral channels associated with the x, y, or z image (recall that z is the stacked image, so $d_z = d_x + d_y$). In the $\nu \to \infty$ limit, the multivariate t becomes Gaussian, $F_\nu(\xi)$ becomes ξ, and Eq. (12) reverts to Eq. (3).

Table 6 compares performance of HACD with the elliptically-contoured variant EC-HACD, proposed in Ref. [11]. Although the experiments in Ref. [11] found an advantage to using EC-HACD over HACD, we observe here that HACD achieves the lowest false alarm rates.

4. CONCLUSIONS

Although we did not embark upon this study with a particular hypothesis in mind, we can make a few observations:

1. The Viareggio 2013 Trial provides a great dataset for change detection. It is in some ways imperfect (e.g., there are changes – actual changes on the ground – that are not part of the ground-truth data) but in other ways this imperfection is useful in that it is representative of the imperfections in real data. The co-registration is imprecise, for instance, but this enables us to evaluate different approaches for mitigating the effects of misregistration. That there are changes in both X→Y and Y→X scenarios enables comparison of symmetric and asymmetric algorithms in both symmetric and asymmetric scenarios.

   Even a great dataset is ultimately anecdotal, however. That one algorithm outperforms another on this dataset is not proof that it is a fundamentally better algorithm, or that it will outperform the other algorithms on other datasets. A good dataset provides a nice counterpoint to theory, but is not a replacement for it.

2. For this data set, local co-registration adjustment (LCRA) makes a big difference. When misregistration is significant (which is more common for airborne imagery than for satellite imagery), LCRA can reduce false alarm rates by a large factor. As expected from the theory, we found that SLCRA worked better in symmetric scenarios, while LCRA was better for asymmetric change detection.

3. We also found that non-maximal suppression (NMS), by discounting false alarms that appear in clumps, improved performance for this data. Further, we find that the NMS improvements are *in addition to* those achieved by LCRA. Comparing Table 7 with Table 1, we find that using LCRA and NMS gives orders of magnitude of improvement in the false alarm rate.

4. CCA is recommended. Reducing dimension from hundreds of channels down to a few tens led in some cases (HACD) to essentially identical performance and in other cases (RX) to substantially better performance. Using fewer channels makes the computation quicker, and would make more complicated nonlinear approaches (e.g., using machine learning) more feasible.

5. Spectral binning and noise whitening provide multiple variants of the Viareggio data, but neither of them had a strong effect on change detection performance.

6. We were surprised by some of the results. We expected HACD to outperform CC for symmetric scenarios, and it did, but we did not expect it to be better for asymmetric scenarios as well. Previous studies indicated that elliptically-contoured variants of CC and HACD would outperform their Gaussian counterparts, but even though this data was measurably fat-tailed, we found that the Gaussian variants of the algorithms worked best.

We only considered a limited number of algorithms – RX, CC, HACD, and their EC counterparts. Future work might include CE and MAD algorithms as well. This dataset may also provide a useful test-bed for more sophisticated algorithms, e.g., using kernels or perhaps more modern machine learning tools.

5. ACKNOWLEDGMENTS

REFERENCES

1. V. Ortiz-Rivera, M. Vélez-Reyes, and B. Roysam, "Change detection in hyperspectral imagery using temporal principal components," *Proc. SPIE* **6233**, p. 623312, 2016.
2. R. Porter, N. R. Harvey, and J. Theiler, "A change detection approach to moving object detection in low frame rate video," *Proc. SPIE* **7341**, p. 73410S, 2009.

3. S. M. Adler-Golden, S. C. Richtsmeier, and R. Shroll, "Suppression of subpixel sensor jitter fluctuations using temporal whitening," *Proc. SPIE* **6969**, p. 69691D, 2008.

4. J. Theiler and S. M. Adler-Golden, "Detection of ephemeral changes in sequences of images," *Proc 37th IEEE Applied Imagery Pattern Recognition Workshop (AIPR)* , 2009.

5. A. Schaum and A. Stocker, "Long-interval chronochrome target detection," *Proc. ISSSR (International Symposium on Spectral Sensing Research)* , 1998.

6. A. Schaum and A. Stocker, "Linear chromodynamics models for hyperspectral target detection," *Proc. IEEE Aerospace Conference* , pp. 1879–1885, 2003.

7. A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF post-processing in multispectral bi-temporal image data: new approaches to change detection studies," *Remote Sensing of the Environment* **64**, pp. 1–19, 1998.

8. J. Theiler, "Quantitative comparison of quadratic covariance-based anomalous change detectors," *Applied Optics* **47**, pp. F12–F26, 2008.

9. N. Acito, M. Diani, G. Corsini, and S. Resta, "Introductory view of anomalous change detection in hyperspectral images within a theoretical Gaussian framework," *IEEE Aerospace and Electronic Systems Magazine* **32**, pp. 2–27, July 2017.

10. A. Ziemann and S. Matteoli, "Detection of large-scale and anomalous changes," in *Hyperspectral Image Analysis: Advances in Machine Learning and Signal Processing*, S. Prasad and J. Chanussot, eds., pp. 351–375, Springer, 2020.

11. J. Theiler, C. Scovel, B. Wohlberg, and B. R. Foy, "Elliptically-contoured distributions for anomalous change detection in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters* **7**, pp. 271–275, 2010.

12. C. Clifton, "Change detection in overhead imagery using neural networks," *Applied Intelligence* **18**, pp. 215–234, 2003.

13. F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Trans. Geoscience and Remote Sensing* **46**(7), pp. 2070–2082, 2008.

14. I. Steinwart, J. Theiler, and D. Llamocca, "Using support vector machines for anomalous change detection," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* , pp. 3732–3735, 2010.

15. J. Theiler and S. Perkins, "Proposed framework for anomalous change detection," *ICML Workshop on Machine Learning Algorithms for Surveillance and Event Detection* , pp. 7–14, 2006.

16. J. Theiler, "Matched-pair machine learning," *Technometrics* **55**(4), pp. 536–547, 2013.

17. J. Theiler, "Transductive and matched-pair machine learning for difficult target detection problems," *Proc. SPIE* **9088**, p. 90880E, 2014.

18. A. Ziemann, M. Kucer, and J. Theiler, "A machine learning approach to hyperspectral detection of solid targets," *Proc. SPIE* **10644**, p. 1064404, 2018.

19. G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. L. Rojo-Alvarez, and M. Martinez-Ramon, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Trans. Geoscience and Remote Sensing* **46**(6), pp. 1822–1835, 2008.

20. M. Volpi, D. Tuia, G. Camps-Valls, and M. Kanevski, "Unsupervised change detection with kernels," *IEEE Geoscience and Remote Sensing Letters* **9**(6), 2012.

21. N. Longbotham and G. Camps-Valls, "A family of kernel anomaly change detectors," in *Proc. 6th IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2014.

22. J. A. Padrón-Hidalgo, V. Laparra, N. Longbotham, and G. Camps-Valls, "Kernel anomalous change detection for remote sensing imagery," *IEEE Trans. Geoscience and Remote Sensing* **57**, pp. 7743–7755, 2019.

23. A. Rossi, N. Acito, M. Diani, G. Corsini, S. Ugo De Ceglie, A. Riccobono, and L. Chiarantini, "Hyperspectral data collection for the assessment of target detection algorithms: the 'Viareggio 2013 Trial'," *Proc. SPIE* **9250**, p. 92500V, 2014.

24. N. Acito, S. Matteoli, A. Rossi, M. Diani, and G. Corsini, "Hyperspectral airborne 'Viareggio 2013 Trial' data collection for detection algorithm assessment," *IEEE J. Selected Topics in Applied Earth Observations and Remote Sensing* **9**, pp. 2365–2376, 2016.

25. S. Matteoli, M. Diani, and G. Corsini, "Target detection experiments with a non-parametric detector on a new hyperspectral data set," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1328–1331, IEEE, 2017.

26. S. Matteoli, M. Diani, and G. Corsini, "Automatic target recognition within anomalous regions of interest in hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **11**(4), pp. 1056–1069, 2018.

27. C. Wu, B. Du, and L. Zhang, "Hyperspectral anomalous change detection based on joint sparse representation," *ISPRS Journal of Photogrammetry and Remote Sensing* **146**, pp. 137–150, 2018.

28. C. Wu, Y. Lin, B. Du, and L. Zhang, "A study for hyperspectral anomaly change detection on 'Viareggio 2013 Trial' dataset," *Proc. 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)* , 2019.

29. J. Theiler, "Subpixel anomalous change detection in remote sensing imagery," *Proc. IEEE Southwest Symposium on Image Analysis and Interpretation* , pp. 165–168, 2008.

30. C. Wu, L. Zhang, and B. Du, "Hyperspectral anomaly change detection with slow feature analysis," *Neurocomputing* **151**, pp. 175–187, 2015.

31. I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoustics, Speech, and Signal Processing* **38**, pp. 1760–1770, 1990.

32. J. Theiler and B. Wohlberg, "Local co-registration adjustment for anomalous change detection," *IEEE Trans. Geoscience and Remote Sensing* **50**, pp. 3107–3116, 2012.

33. K. Vongsy, M. T. Eismann, and M. J. Mendenhall, "Extension of the linear chromodynamics model for spectral change detection in the presence of residual spatial misregistration," *Trans. Geoscience and Remote Sensing* **53**, pp. 3005–3021, 2015.

34. N. Acito, S. Resta, M. Diani, and G. Corsini, "Residual misregistration noise estimation in hyperspectral anomalous change detection," *Optical Engineering* **51**, p. 111705, 2012.

35. M. E. Zelinski, J. R. Henderson, and E. L. Held, "Image registration and change detection for artifact detection in remote sensing imagery," *Proc. SPIE* **10644**, p. 1064413, 2018.